AFRL-AFOSR-VA-TR-2017-0028

Dynamic Decision Making under Uncertainty and Partial Information

Enlu Zhou
GEORGIA TECH RESEARCH CORPORATION
505 10TH ST NW
ATLANTA, GA 30318-5775

01/30/2017
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTA2

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
31-01-2017	Final Report	From 01/02/2014 To 01/31/2017

4. TITLE AND SUBTITLE		5a. CONTRACT NUMBER
Dynamic Decision Making under Uncertainty and Partial Information		
		5b. GRANT NUMBER
		FA- 9550-14-1-0059
		5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Enlu Zhou	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Georgia Institute of Technology H. Milton Stewart School of Industrial & Systems Engineering 755 Ferst Drive, NW, Atlanta, GA 30332	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
Air Force Office of Scientific Research 875 North Randolph Street Suite 325, Room 3112 Arlington VA, 22203	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION A: Distribution approved for public release.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project is concerned with the study of basic questions aimed at meeting challenges in information superiority, logistics, and planning for the Air Force of the future. For successful military operations, the future requirements of the Air Force will include information fusion at a much larger scale and much more agile, responsive, and integrated systems. Such problems and systems are exceedingly complex; however, a central part of them is decision making, which often takes place sequentially in time, subject to uncertainty in the future and limited partial information at hand. In order to address these problems, we investigated efficient computational methodologies for dynamic decision making under uncertainty and partial information. In the course of this research, we developed and studied efficient simulation-based methodologies for dynamic decision making under uncertainty and partial information; (ii) studied the application of these decision making models and methodologies to practical problems, such as those arising in planning, logistics, and risk management. The proposed research resulted in (i) new mathematical tools and theories for dynamic decision making and optimal stopping; (ii) useful application of these models and new methodologies in a wide range of problems.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
					19b. TELEPHONE NUMBER (Include area code)

Reset

Standard Form 298 (Rev. 8/98)

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Dynamic Decision Making under Uncertainty and Partial Information
Final Performance Report
Grant FA-9550-14-1-0059
Funding period: 02/01/2014 - 01/31/2017
Report date: January 29, 2017

Principal Investigator: Enlu Zhou

H. Milton Stewart School of Industrial & Systems Engineering

Georgia Institute of Technology

Email: enlu.zhou@isye.gatech.edu; Phone: 404-385-1581

Abstract

The researchers made significant progress in all of the proposed research areas. The first major task in the proposal involved duality in stochastic control and optimal stopping. In support of this task, the researchers developed new methods for efficiently solving optimal stopping problems of partially observable Markov processes and optimal stopping problems under jump-diffusion processes. The researchers also studied the information relaxation approach and established duality for controlled Markov diffusions and weakly coupled dynamic programs. In the second major task aiming at solving difficult global optimization problems, the researchers proposed and developed a new framework that integrates the idea of model-based randomized optimization with gradient-based optimization, and further extended this method to simulation optimization problems. All the developed methods have been tested through numerical experiments and demonstrated excellent performance. The methods have also been applied to problems in revenue management, option pricing, and power allocation in communication networks.

1 Introduction

In this research project, we carried out research concerning with the study of basic questions aimed at meeting challenges in information superiority, logistics, and planning for the Air Force of the future. For successful military operations, the future requirements of the Air Force will include information fusion at a much larger scale and much more agile, responsive, and integrated systems. Such problems and systems are exceedingly complex; however, a central part of them is decision making, which often takes place sequentially in time, subject to uncertainty in the future and limited partial information at hand. In order to address these problems, we investigated efficient computational methodologies for dynamic decision making under uncertainty and partial information. In the course of this research, we proposed to (i) develop and study efficient simulation-based methodologies for dynamic decision making under uncertainty and partial information; (ii) study the application of these decision making models and methodologies to practical problems, such as those arising in planning, logistics, and risk management. The completed research resulted in (i) new mathematical tools and theories for dynamic decision making and optimal stopping; (ii) powerful algorithms for solving problems that currently cannot be tackled by existing methods; (iii) useful application of these models and new methodologies in a wide range of problems. Our work can be effective tools in an integrated approach to Global Awareness (Intelligence, Surveillance and Reconnaissance, or ISR), Command and Control (C2), planning, and logistics for the Air Force.

In particular, we combined three approaches in our research:

- developing efficient simulation-based methodologies that have a sound theoretical basis;

- providing convergence guarantees and error bounds for the developed algorithms;

- studying the application of these decision making models and methodologies to practical problems, such as those arising in revenue management, risk, and communication networks.

1.1 Duality in Stochastic Control and Optimal Stopping

Stochastic control and optimal stopping provide a powerful paradigm for modelling sequential optimization under uncertainty. The typical models include Markov decision processes (MDPs), controlled Markov diffusion, optimal stopping under full or partial observation. To be more specific about the problems we have considered, we describe the MDP below, and the other models are formulated similarly. Consider a finite-horizon MDP on the probability space $(\Omega, \mathcal{G}, \mathbb{P})$. Time is indexed by $\mathcal{K} = \{0, 1, \cdots, K\}$. Suppose $\mathcal{X}$ is the state space and $\mathcal{A}$ is the action space. The state $\{x_k\}$ follows the equation

$$x_{k+1} = f(x_k, a_k, v_{k+1}), \ k = 0, 1, \cdots, K - 1, \tag{1}$$

where $a_k \in \mathcal{A}_k$ is the action whose value is decided at time k, and v_k is a random variable with a known distribution taking values in the set $\mathcal{V}$. The evolution of the information is described by the filtration $\mathbb{G} = \{\mathcal{G}_0, \cdots, \mathcal{G}_K\}$ with $\mathcal{G} = \mathcal{G}_K$. In particular, each v_k is $\mathcal{G}_k$-adapted. Denote by $\mathbb{A}$ the set of all policies $\mathbf{a} \triangleq (a_0, \cdots, a_{K-1})$, i.e., each a_k takes value in $\mathcal{A}$. Let $\mathbb{A}_\mathbb{G}$ be the set of *non-anticipative* policies that are adapted to the filtration $\mathbb{G}$, i.e., each a_k is $\mathcal{G}_k$-adapted. Given an $x_0 \in \mathcal{X}$, the objective is to maximize the expected reward by selecting a non-anticipative policy $\mathbf{a} \in \mathbb{A}_\mathbb{G}$:

$$V_0(x_0) = \sup_{\mathbf{a} \in \mathbb{A}_\mathbb{G}} \mathbb{E}\left[\sum_{k=0}^{K-1} g_k(x_k, a_k) + \Lambda(x_K)|x_0\right]. \tag{2}$$

A-2

The expectation in (2) is taken with respect to the random sequence $\mathbf{v} = (v_1, \cdots, v_K)$.

Numerical solutions to stochastic control and optimal stopping problems often suffer from the so-called "curse of dimensionality", i.e., the computational complexity grows exponentially as the dimension of the state increases. To overcome this difficulty, a variety of approximate dynamic programming techniques have been developed, see [6, 25, 3]. These methods often generate sub-optimal policies, which lead to lower bounds on the optimal expected reward by simulating the dynamic system under the aforementioned policies. Though the accuracy of the sub-optimal policies is generally unknown, the lack of performance guarantee on sub-optimal policies can be potentially addressed by providing an upper bound on the optimal expected reward. Towards this end, recently researchers have spent a significant amount of effort in studying the duality in stochastic control and optimal stopping. The main idea of this dual approach is to allow the decision maker to foresee the future uncertainty but impose a penalty for getting access to the information in advance. The optimal expected reward can be achieved by imposing an optimal penalty, which is not directly available in practice. Hence, approximation schemes based on different types of optimal penalties have been proposed in order to derive tight dual bounds, such as [27] and [5]..

The PI and her collaborators' contribution includes: (i) a filtering-based duality approach for solving optimal stopping problems under partial observation; (ii) a parameterized duality approach for solving MDPs; (iii) development of duality theory in controlled Markov diffusions; (iv) an approach combining information relaxation with Lagrangian relaxation to solve large-scale weakly coupled dynamic programs; (v) highly efficient algorithm for solving dual problem of dynamic programming.

1.2   Model-based Global Optimization

In dynamic decision making and control, one-stage optimization often comes up as a subproblem. Hence, we considered the following optimization problem

$$x^* \in \arg\max_{x \in \mathcal{X}} H(x), \quad \mathcal{X} \subseteq \mathbb{R}^n. \tag{3}$$

where the solution space $\mathcal{X}$ is a nonempty compact set in $\mathbb{R}^n$, and $H : \mathcal{X} \to \mathbb{R}$ is a real-valued function. Denote the optimal function value as H^*, i.e., there exists an x^* such that $H(x) \leq H^* \triangleq H(x^*)$, $\forall x \in \mathcal{X}$. Assume that H is bounded on $\mathcal{X}$, i.e., $\exists H_{lb} > -\infty$, $H_{ub} < \infty$ s.t. $H_{lb} < H(x) < H_{ub}$, $\forall x \in \mathcal{X}$.

Model-based global optimization methods use probability distributions to weight promising areas of the solution space, where the distribution is updated iteratively based on output from the samples drawn according to the current distribution. They are well suited for global optimization problems where there is limited structural information on the optimizing function (e.g., derivatives, convexity).

The PI and her collaborators' contribution includes: (i) the introduction of the new idea of incorporating model-based optimization into direct gradient search; (ii) development of a rigorous algorithm framework of gradient-based adaptive stochastic search; (iii) development of a series algorithms for black-box optimization, simulation optimization problems, and release of user-friendly software.

2   Research Results

2.1   Duality in Stochastic Control and Optimal Stopping

Building upon the recently-developed duality theory in stochastic dynamic programs, we tackled stochastic control and optimal stopping problems (either with full observation or partial observation) from three angles: (i) we developed the duality theory and proposed a parameterized penalty framework in the dual formulation based on the new insights we gained into the structure of optimal penalty functions; (ii) we incorporated advanced Monte Carlo techniques to develop highly efficient algorithms in evaluating the dual formulation; (iii) we studied the numerical performance of our algorithms on a wide range of application problems in revenue and risk management.

In [31][33] we focused on optimal stopping under partial observation and developed a lower-and-upper-bound approach with moderate computational cost. The motivation is that the gap between the lower and upper bounds gives an indication of the quality of the approximate solutions. To guarantee a high-quality approximate solution, we can increase the computation effort until the gap between the two bounds decreases to a desirable tolerance level. We proposed a filtering-based duality approach that complements a suboptimal stopping time (hence an asymptotic lower bound) with an asymptotic upper bound on the value function. Since our approach does not tie to a particular model and only involves Monte Carlo simulation, it can be generalized to any POMP as long as the particle filtering technique can be applied. Our method relies on the martingale duality formulation of the fully observable optimal stopping problem, which is proposed by [26] and [13] in the setting of pricing American options under constant volatility. The numerical method in [13] generates the suboptimal martingale based on approximate value functions; whereas later [1] developed an alternative duality-based method that uses a suboptimal policy to generate the martingale. From the perspective of modeling fidelity versus computational complexity, it is not trivial to compare optimal stopping of POMPs with its counterpart in fully observable Markov processes. In particular, the difference of their value functions cannot be quantified in general and is problem dependent, so we were also interested in learning the features that influence this difference in the underlying probabilistic model. Indeed, as an example, our numerical experiments on pricing American options under partially observable stochastic volatility showed that our asymptotic upper bound is strictly less than the option price of the model where the volatility is treated directly observable, and the difference is especially obvious when the effect of the volatility is dominant. This in turn showed that our method provides a better criterion to evaluate the performance of a suboptimal policy in the partially observable model.

In [32], we noted that the construction of a good penalty in the dual formulation of MDPs is usually difficult due to the tradeoff between its effectiveness and the computational cost. As a consequence, all penalties developed so far for continuous-state MDPs are linear functions for the sake of maintaining the solvability of the intermediate pathwise optimization problem. To capture the nonlinear feature of the optimal penalty in a general case, we introduced a class of simple nonlinear penalty functions that can be applied to general MDPs. This class of nonlinear penalties together with other classes of linear penalties developed in [4] and [10], all lead to dual bounds on the value function. In summary, our contributions are: (i) we developed a framework of parameterized penalties in the dual representation of MDPs, where the optimal choice of the parameters can be determined by a convex (stochastic) optimization problem. The theoretic result guarantees a tighter dual bound if more penalties are used; (ii) we introduced a new class of nonlinear penalties that can be applied to general MDPs and are also very easy to implement in practice; (iii) we carried out some numerical experiments that provide insights into the design and choice of penalties. The numerical results showed a considerable improvement on the tightness of

the dual bound using our parameterized penalties.

In [34], we extended the information relaxation approach and the dual representation of MDPs to controlled Markov diffusions. The motivation is that the Hamilton-Jacobi-Bellman (HJB) equation rarely allows a closed-form value function, especially when the dimension of the state space is high or there are constraints on the control space. Many numerical methods have been developed based on different approximation schemes: [16] considered the Markov chain approximation method by discretizing the HJB equation; [12] extended the approximate linear programming method to controlled Markov diffusions. Another standard numerical approach is to discretize the time space, which reduces the original continuous-time problem to an MDP and hence the techniques of approximate dynamic programming can be applied. Since the quality of a numerical solution is hard to justify in many problems, we were interested in deriving a tight dual bound on the value function of a controlled Markov diffusion by formulating its dual representation. To see if it is possible to establish a similar framework of dual formulation for controlled Markov diffusions based on information relaxation as that for MDPs, we presented the information relaxation-based dual formulation of controlled Markov diffusions based on the technical machinery "anticipating stochastic calculus" (see, e.g., [24, 23]). We further established the weak duality, strong duality and complementary slackness results in a parallel way as those in the dual formulation of MDPs. We investigated one type of optimal penalties, i.e., the so-called "value function-based penalty", which has the key feature that it can be written compactly as an Ito stochastic integral under the natural filtration generated by the Brownian motions. This compact expression potentially enables us to design sub-optimal penalties in simple forms and also facilitates the computation of the dual bound. Then we emphasized on the computational aspect using the value function-based optimal penalty so as to answer the third question. A direct application is illustrated by a classic dynamic portfolio choice problem with predictable returns and intermediate consumptions: we considered the numerical solution to a discrete-time model that is discretized from a continuous-time model; an effective class of penalties that are easy to evaluate is proposed to derive dual bounds on the value function for the discrete-time model. In addition, we found that the Lagrangian approach proposed early by [9] has a similar flavor as the gradient-based penalty proposed by [4] for MDPs. The main difference of their work compared with ours is that we propose a more general framework that may incorporate their Lagrangian approach as a special case; the optimal penalty we developed in this work is value function-based while their Lagrangian approach behaves like a gradient-based penalty. In summary, our contributions include: (i) we established a dual representation of controlled Markov diffusions based on information relaxation; (ii) we also explored the structure of the optimal penalty and exposed the connection between MDPs and controlled Markov diffusions; (iii) based on the result of the dual representation of controlled Markov diffusions, we demonstrated its practical use in a dynamic portfolio choice problem. In many cases the numerical results of the upper bounds on the expected utility show that our proposed penalties are near optimal, comparing with the lower bounds induced by sub-optimal policies for the same problem.

In [40][39], we considered pricing American-style derivatives, which is essentially an optimal stopping problem. There has been an active and challenging problem in the last thirty years, especially when the underlying stocks' prices follow some jump-diffusion processes, as they become more and more critical to investors. To present time, various jump-diffusion models for financial modelling have been proposed to fit the real data in financial markets, including the normal jump-diffusion model, the affine jump-diffusion models, the jump models based on Levy processes, the double exponential, mixed-exponential and hyper-exponential jump-diffusion models. All these models are trying to capture some interesting features of the market behaviour that cannot be well explained by pure-diffusion models, such as the heavy-tail risk suffered by the market. In general, closed-form expressions for the American-style derivatives can hardly be derived under

these jump-diffusion models due to the multiple exercise opportunities and the randomness in the underlying asset price caused by both jumps and diffusions. Hence, various numerical methods have been proposed to tackle the American-style option pricing problems under the jump-diffusion models. We generalized the idea of "true martingale" in [2] to Bermudan option pricing problem. We proved that the true martingale approximation converges to the objective martingale in the mean square sense provided that the time discretization goes to zero by bounding the empirical difference between the approximation and the objective martingale. In numerical experiments, we investigated the numerical effectiveness of the least-squares regression approach (L-S algorithm) [19] for Bermudan option price under the jump-diffusion models. In particular, we found that by incorporating the European option price under the corresponding pure-diffusion model (referred as the "non-jump European option") into the function basis of the L-S algorithm, the quality of the induced suboptimal exercise strategies and the lower bounds can be significantly improved. Motivated by the explicit structure of the optimal dual martingale, we proposed a function basis that can be employed in the T-M algorithm to obtain tight upper bounds on the option price. By implementing our algorithm and the A-B algorithm on several sets of numerical experiments, the numerical results demonstrate that both methods can generate tight and stable upper bounds on option price; however, we observed that our algorithm is much more efficient than the A-B algorithm in practice due to the relief from nested simulation.

In [35], we studied the weakly coupled dynamic program, which describes a broad class of stochastic optimization problems in which multiple controlled stochastic processes evolve independently but subject to a set of linking constraints imposed on the controls. One feature of the weakly coupled dynamic program is that it decouples into lower dimensional dynamic programs by dualizing the linking constraint via the Lagrangian relaxation, which yields a bound on the optimal value of the original dynamic program. Together with the Lagrangian bound, we generalized the information relaxation approach that relaxes the non-anticipative constraint on the controls to obtain a tighter dual bound. To tackle large-scale problems, we further proposed a computationally tractable method based on information relaxation, and showed it provides a valid dual bound and its performance has a uniform bound regardless of the number of subproblems. We implemented our method and demonstrate its use on a dynamic product promotion problem and a linear quadratic control problem with nonconvex linking constraints.

In [41], we developed a framework of regression approach to approximating the optimal dual penalty in a non-nested manner, by exploring the structure of the function space consisting of all feasible dual penalties. The resulted approximations maintain to be feasible dual penalties, and thus yield valid dual bounds on the optimal value function. We showed that the proposed framework is computationally efficient, and the resulted dual penalties lead to numerically tractable dual problems. Finally, we applied the framework to a high-dimensional dynamic trading problem to demonstrate its effectiveness in solving the dual problems of complex dynamic programs.

2.2   Gradient-based Adaptive Stochastic Search

We distinguish between instance-based and model-based global optimization solution methods. In instance-based methods, the search for new candidate solutions depends directly on previously generated solutions, e.g., simulated annealing, genetic algorithms (GAs), tabu search, and nested partitions. On the other hand, model-based algorithms typically assume a sampling distribution (i.e., a probabilistic model), often within a parameterized family of distributions, over the solution space, and iteratively carry out the two interrelated steps: (1) draw candidate solutions from the sampling distribution; (2) use the evaluations of these candidate solutions to update the sampling distribution. The hope is that at every iteration the sampling distribution is biased towards the

more promising regions of the solution space, and will eventually concentrate on one or more of the optimal solutions. Examples of model-based algorithms include ant colony optimization [11], annealing adaptive search (AAS) [28], probability collectives (PCs) [30], the estimation of distribution algorithms (EDAs) [17], the cross-entropy (CE) method [29], model reference adaptive search (MRAS) [14], and the interacting-particle algorithm [21, 22]. The various model-based algorithms mainly differ in their ways of updating the sampling distribution. Because model-based algorithms work with a population of candidate solutions at each iteration, they demonstrate more robustness in exploring the solution space as compared with their classical counterparts that work with a single candidate solution each time (e.g., simulated annealing). have found widespread applications in solving hard nonlinear optimization problems and have the potential of being useful tools in a number of areas related to estimation and control [6, 8, 18, 20].

In [36][38], we proposed a gradient-based adaptive stochastic search (GASS) framework for solving general optimization problems with little structure. The main motivation is to integrate this robustness feature of model-based algorithms into familiar gradient-based tools from classical differentiable optimization to facilitate the search for good sampling distributions. The underlying idea is to reformulate the original (possibly non-differentiable) optimization problem into a *differentiable* optimization problem over the parameter space of the sampling distribution, and then use a direct gradient search method on the parameter space to solve the new formulation. This leads to a natural algorithmic framework that combines the advantages of both methods: the fast convergence of gradient-based methods and the global exploration of stochastic search. Specifically, each iteration of our proposed method consists of the following two steps: (1) generate candidate solutions from the current sampling distribution; (2) update the parameters of the sampling distribution using a direct gradient search method. Although there are a variety of gradient-based algorithms that are applicable in step (2) above, we focused on a particular algorithm that uses a Newton-like procedure to update the sampling distribution parameters. Note that since the algorithm uses only the information contained in the sampled solutions, it differs from the Newton-like method in deterministic optimization, in that there is an extra Monte Carlo sampling noise involved at each parameter updating step. We showed that this stochastic version of Newton-like iteration can be expressed in the form of a generalized Robbins-Monro algorithm, and this in turn allowed us to use the existing tools from stochastic approximation theory to analyze the asymptotic convergence and convergence rate of the proposed algorithm. The algorithm iteratively finds high quality solutions by randomly sampling candidate solutions from a parameterized distribution model over the solution space, combining the robustness feature of stochastic search from considering a population of candidate solutions with the relative fast convergence speed of classical gradient methods by exploiting local differentiable structures. We analyzed the convergence and converge rate properties of the proposed algorithm, and carried out numerical study to illustrate its performance. In summary, the major contributions of this work includes (1) establish an algorithm framework that integrates the central idea of model-based optimization with direct gradient search; (2) develop a series of algorithms for non-differentiable or even black-box optimization.

In [7], under the framework of GASS, we proposed two discrete optimization algorithms: discrete gradient-based adaptive stochastic search (discrete-GASS) and annealing gradient-based adaptive stochastic search (annealing-GASS). In discrete-GASS, we transformed the discrete optimization problem to a continuous optimization problem on the parameter space of a family of independent discrete distributions, and applied a gradient-based method to find the optimal parameter such that the corresponding distribution has the best capability to generate optimal solution(s) to the original discrete problem. In annealing-GASS, we used Boltzmann distribution as the parameterized probabilistic model, and propose a gradient-based temperature schedule which changes adaptively with respect to the current performance of the algorithm. We proved the convergence of the two

proposed methods, and conduct numerical experiments to compare these two methods as well as some other existing methods. More specifically, the first algorithm (discrete-GASS) uses the independent discrete distribution as the parameterized distribution, where the parameters determine the probabilities that the components of a candidate solution will be selected. By introducing this probabilistic model, instead of directly solving the solutions on the discrete space, we converted the discrete optimization problem to a continuous problem on the parameter space, and hence we can apply a gradient-based method on the parameter space to find the optimal probabilistic model. In this algorithm, we can easily generate samples according to the parameterized distribution; however, the dimensionality of the transformed problem on the parameter space is large compared to the original problem. Therefore, we proposed the second algorithm (annealing-GASS), which uses the Boltzmann distribution as the probabilistic model that only has a scalar parameter known as the temperature. The dimension of the parameter is only one regardless of the dimension or cardinality of the solution space. Different from discrete-GASS, where the optimal parameter is unknown and needs to be solved by a gradient-based method, the optimal parameter (temperature) in annealing-GASS is actually well known to be zero; however, generating samples exactly from a sequence of Boltzmann distributions with time-dependent temperatures and choosing an appropriate temperature schedule are difficult and challenging tasks in the implementation. Various choices of temperature schedules have been studied and tested in the literature [12, 23, 27, 5, 9, 25]. We derived a temperature schedule under the GASS framework by using a gradient-based method to solve the reformulated objective function on the parameter and adaptively updating the parameter based on the current performance of the algorithm. In summary, the major contributions of this work includes (1) extension of gradient-based adaptive stochastic search to discrete optimization with convergence results; (2) a new adaptive temperature schedule for a converging sequence of Boltzmann distributions.

In [15], we aimed to improve the efficiency of model-based algorithms by reducing the number of candidate solutions generated per iteration. This was carried out through embedding a stochastic averaging procedure within these methods to make more efficient use of the past sampling information. This procedure not only can potentially reduce the number of function evaluations needed to obtain high-quality solutions, but also makes the underlying algorithms more amenable for parallel computation. The detailed implementation of our approach is demonstrated through an exemplary algorithm instantiation called Model-based Annealing Random Search with Stochastic Averaging (MARS-SA), which maintains the per iteration sample size at a small constant value. We established the global convergence property of MARS-SA and provided numerical examples to illustrate its performance.

In [37], we extended the idea of model-based algorithms for deterministic optimization to simulation optimization over continuous space. Model-based algorithms iteratively generate a population of candidate solutions from a sampling distribution and use the performance of the candidate solutions to update the sampling distribution. By viewing the original simulation optimization problem as another optimization problem over the parameter space of the sampling distribution, we proposed to use a direct gradient search on the parameter space to update the sampling distribution. To improve the computational efficiency, we further developed a two-timescale updating scheme that updates the parameter on a slow timescale and estimates the quantities involved in the parameter updating on the fast timescale. We analyzed the convergence properties of our algorithms through techniques from stochastic approximation, and illustrated the performance of our algorithms by comparing with two state-of-the-art model-based simulation optimization methods.

3 Research Output from AFOSR support

3.1 Publications

1. Enlu Zhou* and Shalabh Bhatnagar, "Gradient-based Adaptive Stochastic Search for Simulation Optimization over Continuous Space", under 3rd round review, INFORMS Journal on Computing.

2. Helin Zhu, Fan Ye, and Enlu Zhou*, A Regression Approach to Dual Problems in Dynamic Programming", conditionally accepted, IEEE Transactions on Automatic Control.

3. Joshua Hale, Enlu Zhou*, and Jiming Peng, "A Lagrangian Search Method for the P-Median Problem", to appear, Journal of Global Optimization, 2016.

4. Xi Chen and Enlu Zhou*, Population Model-Based Optimization, Journal of Global Optimization, 63(1), pp. 125-148, 2015.

5. Fan Ye and Enlu Zhou*, Duality and Information Relaxation in Controlled Markov Diffusions, IEEE Transactions on Automatic Control, 60(10), pp. 2676-2691, 2015.

6. Helin Zhu, Fan Ye, and Enlu Zhou*, Fast Estimation of True Bounds on Bermudan Option Price under Jump Diffusion Processes, Quantitative Finance, 15(11), pp. 1885-1900, 2015.

7. Jiaqiao Hu, Enlu Zhou*, and Qi Fan, Model-based Optimization Annealing Random Search with Stochastic Averaging, ACM Transactions on Modeling and Computer Simulation, 24(4), pp. 21:1-21:23, 2014.

8. Enlu Zhou* and Jiaqiao Hu, Gradient-based Adaptive Stochastic Search for Non-differentiable Optimization, IEEE Transactions of Automatic Control, 59(7), pp. 1818-1832, 2014.

9. Fan Ye and Enlu Zhou*, Optimal Stopping of Partially Observable Markov Processes: A Filtering-based Duality Approach, IEEE Transactions on Automatic Control, 58(10), pp. 2698-2704, 2013.

10. Enlu Zhou*, Optimal Stopping under Partial Observation: Near-Value Iteration, IEEE Transactions on Automatic Control, 58(2), pp. 500-506, 2013.

11. Enlu Zhou*, Shalabh Bhatnagar, Xi Chen, "Simulation Optimization via Gradient-based Stochastic Search", in Proceedings of the 2014 Winter Simulation Conference, pp. 3869-3879, 2014.

12. Fan Ye and Enlu Zhou*, Dual Formulation of Controlled Markov Diffusions and Its Application", in Proceedings of the 19th IFAC World Congress, pp. 7811-7818, 2014.

3.2 Software

We developed a user-friendly software based on our algorithms developed under GASS. There are two versions of the software: one is an application in Matlab, and the other is a standalone application. Both are available for free download on the PI's website http://enluzhou.gatech.edu/software.html. The website also includes user manual, demos, and all the Matlab code. This software provides an interactive environment to solve deterministic optimization problems with multiple local optima,

problems with little structural properties, and black-box optimization problems, using the GASS method. A list of algorithms provided for different types of problems is shown below.

	Continuous Problem	Discrete Problem
Unconstraint	Continuous-GASS Continuous-GASS with Averaging	Discrete Annealing-GASS
Linear Equality Constraint	Continuous-GASS (Linear Equality Constraint)	
Linear Inequality Constraint	Continuous-GASS (Linear Inequality Constraint)	
Box Constraint		Discrete -GASS

The main user interface of the software is shown below.
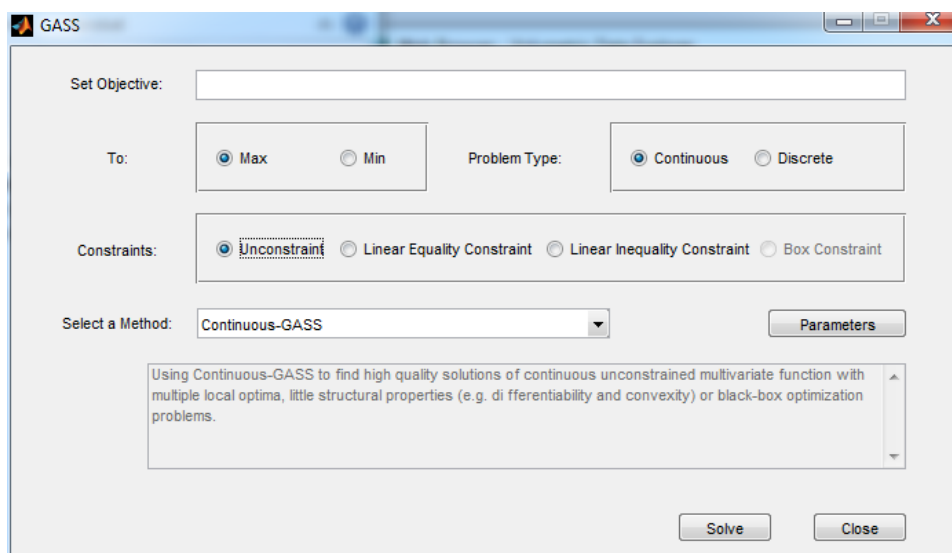
3.3 Students Supported

1. Fan Ye, Ph.D. graduated in 2015, Georgia Institute of Technology, now Associate at Morgan Stanley.

2. Helin Zhu, Ph.D. graduated in 2016, Georgia Institute of Technology, now data scientist at Uber.

3. Joshua Hale, Ph.D. expected in 2017, Georgia Institute of Technology.

References

[1] L. Andersen and M. Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222 – 1234, 2004.

[2] D. Belomestny, C. Bender, and J. Schoenmakers. True upper bounds for Bermudan products via non-nested Monte Carlo. *Mathematical Finance*, 19:53 – 71, 2009.

[3] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 4th edition, 2012.

[4] D.B. Brown and J.E. Smith. Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds. *Management Science*, 57(10):1752–1770, 2011.

[5] D.B. Brown, J.E. Smith, and P. Sun. Information relaxations and duality in stochastic dynamic programs. *Operations Research*, 58(4):758 – 801, 2010.

[6] H. S. Chang, M. C. Fu, J. Hu, and S. I. Marcus. *Simulation-based Algorithms for Markov Decision Processes*. Communications and Control Engineering Series. Springer, New York, 1st edition, 2007.

[7] Xi Chen, Enlu Zhou, and Jiaqiao Hu. Discrete optimization via gradient-based adaptive stochastic search methods. *IIE Transactions*, 2014. Under revision.

[8] F. Dambreville. Cross-entropic learning of a machine for the decision in a partially observable universe. *Journal of Global Optimization*, 37:541–555, 2007.

[9] M.H.A. Davis and G. Burstein. A deterministic approach to stochastic optimal control with application to anticipative control. *Stochastics: An International Journal of Probability and Stochastic Processes*, 40(3-4):203–256, 1992.

[10] V. Desai, V. Farias, and C. Moallemi. Bounds for Markov decision processes. Chapter in Reinforcement Learning and Approximate Dynamic Programming for Feedback Control (F. L. Lewis, D. Liu, eds.), 2011.

[11] M. Dorigo and L.M. Gambardella. Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1:53 – 66, 1997.

[12] J. Han and B. van Roy. Control of diffusions via linear programming. *Stochastic Programming*, pages 329–353, 2011.

[13] M.B. Haugh and L. Kogan. Pricing American options: A duality approach. *Operations Research*, 52(2):258 – 270, 2004.

[14] J. Hu, M. C. Fu, and S. I. Marcus. A model reference adaptive search method for global optimization. *Operations Research*, 55:549–568, 2007.

[15] Jiaqiao Hu, Enlu Zhou, and Qi Fan. Model-based annealing random search with stochastic averaging. *ACM Transactions on Modeling and Computer Simulation*, 24(4):21:1–21:23, 2014.

[16] H. J. Kushner and A. S. Budhiraja. A nonlinear filtering algorithm based on an approximation of the conditional distribution. *IEEE Transactions on Automatic Control*, 45(3):580 – 585, 2000.

[17] P. Larranaga, R. Etxeberria, J. A. Lozano, and J. M. Pena. Optimization by learning and simulation of Bayesian and Gaussian networks. Technical Report EHU-KZAA-IK-4/99, Department of Computer Science and Artificial Intelligence, University of the Basque Country, 1999.

[18] F. Liang. Annealing stochastic approximation monte carlo for neural network training. *Machine Learning*, 68:201–233, 2007.

[19] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies*, 14(1):113 – 147, 2001.

[20] S. Mannor, R.Y. Rubinstein, and Y. Gat. The cross-entropy method for fast policy search. In *Proceedings of the 20th International Conference on Machine Learning*, 2003.

[21] O. Molvalioglu, Z. B. Zabinsky, and W. Kohn. The interacting-particle algorithm with dynamic heating and cooling. *Journal of Global Optimization*, 43:329–356, 2009.

[22] O. Molvalioglu, Z. B. Zabinsky, and W. Kohn. Meta-control of an interacting-paricle algorithm. *Nonlinear Analysis: Hybrid Systems*, 4(4):659 – 671, 2010.

[23] D. Nualart. *The Malliavin calculus and related topics.* Springer-Verlag, 2nd edition, 2006.

[24] D. Ocone and E. Pardoux. A generalized it-ventzell formula. application to a class of anticipating stochastic differential equations. *Annales de l'institut Henri Poincar (B) Probabilits et Statistiques*, 25(1):39–71, 1989.

[25] W.B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality.* John Wiley and Sons, 2nd edition, 2011.

[26] L. C. G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271 – 286, 2002.

[27] L. C. G. Rogers. Pathwise stochastic optimal control. *SIAM J.Control Optimization*, 46(3):1116 – 1132, 2007.

[28] H. E. Romeijn and R. L. Smith. Simulated annealing for constrained global optimization. *Journal of Global Optimization*, 5(2):101–126, 1994.

[29] R. Y. Rubinstein. Combinatorial optimization, ants and rare events. In S. Uryasev and P.M. Pardalos, editors, *Stochastic Optimization: Algorithms and Applications*, pages 304–358, Dordrecht, The Netherlands, 2001. Kluwer Academic Publishers.

[30] D. H. Wolpert. Finding bounded rational equilibria part i: Iterative focusing. In T. Vincent, editor, *Proceedings of the Eleventh International Symposium on Dynamic Games and Applications*, 2004.

[31] F. Ye and E. Zhou. Pricing american options under partial observation of stochastic volatility. In *Proceedings of the 2011 Winter Simulation Conference*, pages 3760–3771, 2011.

[32] F. Ye and E. Zhou. Parameterized penalties in the dual representation of markov decision processes. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 870–876, 2012.

[33] Fan Ye and Enlu Zhou. Optimal stopping of partially observable markov processes: A filtering-based duality approach. *IEEE Transactions on Automatic Control*, 58(10):2698–2704, 2013.

[34] Fan Ye and Enlu Zhou. Dual formulation of controlled markov diffusions and its application. In *Proceedings of 19th IFAC World Congres*, 2014.

[35] Fan Ye, Helin Zhu, and Enlu Zhou. Weakly coupled dynamic program: Information and lagrangian relaxations. *IEEE Transcations on Automatic Control*, 2014. Under review.

[36] E. Zhou and J. Hu. Combining gradient-based optimization with stochastic search. In *Proceedings of the 2012 Winter Simulation Conference*, 2012.

[37] Enlu Zhou and Shalabh Bhatnagar. Gradient-based adaptive stochastic search for simulation optimization over continuous space. *INFORMS Journal on Computing*, 2014. Under review.

[38] Enlu Zhou and Jiaqiao Hu. Gradient-based adaptive stochastic search for non-differentiable optimization. *IEEE Transactions on Automatic Control*, 59(7):1818–1832, 2014. Longer version available at http://arxiv.org/abs/1301.1333.

[39] Helin Zhu, Fan Ye, and Enlu Zhou. True upper bounds on bermudan option prices under jump-diffusion processes. In *Proceedings of the 2013 Winter Simulation Conference*, 2013.

[40] Helin Zhu, Fan Ye, and Enlu Zhou. Fast estimation of true bounds on bermudan option price under jump diffusion processes. *Quantitative Finance*, 15(11):1885–1900, 2015.

[41] Helin Zhu, Fan Ye, and Enlu Zhou. A regression approach to dual problems in dynamic programming. *IEEE Transactions on Automatic Control*, 2015. Conditionally accepted.

1.

Report Type

Final Report

Primary Contact Email
Contact email if there is a problem with the report.

enlu.zhou@isye.gatech.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

404-385-1581

Organization / Institution name

Georgia Institute of technology

Grant/Contract Title
The full title of the funded effort.

Dynamic Decision Making under Uncertainty and Partial Information

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-14-1-0059

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Enlu Zhou

Program Officer
The AFOSR Program Officer currently assigned to the award

Jean-Luc Cambier

Reporting Period Start Date

02/01/2014

Reporting Period End Date

01/31/2017

Abstract

This project is concerned with the study of basic questions aimed at meeting challenges in information superiority, logistics, and planning for the Air Force of the future. For successful military operations, the future requirements of the Air Force will include information fusion at a much larger scale and much more agile, responsive, and integrated systems. Such problems and systems are exceedingly complex; however, a central part of them is decision making, which often takes place sequentially in time, subject to uncertainty in the future and limited partial information at hand. In order to address these problems, we investigated efficient computational methodologies for dynamic decision making under uncertainty and partial information. In the course of this research, we developed and studied efficient simulation-based methodologies for dynamic decision making under uncertainty and partial information; (ii) studied the application of these decision making models and methodologies to practical problems, such as those arising in planning, logistics, and risk management. The proposed research resulted in (i) new mathematical tools and theories for dynamic decision making and optimal stopping; (ii) useful application of these models and new methodologies in a wide range of problems.

Distribution Statement

DISTRIBUTION A: Distribution approved for public release.

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

sf0298-2017.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

final+report_2017.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

1. Enlu Zhou* and Shalabh Bhatnagar, "Gradient-based Adaptive Stochastic Search for Simulation Optimization over Continuous Space", under 3rd round review, INFORMS Journal on Computing.

2. Helin Zhu, Fan Ye, and Enlu Zhou*, A Regression Approach to Dual Problems in Dynamic Programming", conditionally accepted, IEEE Transactions on Automatic Control.

3. Joshua Hale, Enlu Zhou*, and Jiming Peng, "A Lagrangian Search Method for the P-Median Problem", to appear, Journal of Global Optimization, 2016.

4. Xi Chen and Enlu Zhou*, Population Model-Based Optimization, Journal of Global Optimization, 63(1), pp. 125-148, 2015.

5. Fan Ye and Enlu Zhou*, Duality and Information Relaxation in Controlled Markov Diffusions, IEEE Transactions on Automatic Control, 60(10), pp. 2676-2691, 2015.

6. Helin Zhu, Fan Ye, and Enlu Zhou*, Fast Estimation of True Bounds on Bermudan Option Price under Jump Diffusion Processes, Quantitative Finance, 15(11), pp. 1885-1900, 2015.

7. Jiaqiao Hu, Enlu Zhou*, and Qi Fan, Model-based Optimization Annealing Random Search with Stochastic Averaging, ACM Transactions on Modeling and Computer Simulation, 24(4), pp.21:1-21:23, 2014.

8. Enlu Zhou* and Jiaqiao Hu, Gradient-based Adaptive Stochastic Search for Non-differentiable Optimization, IEEE Transactions of Automatic Control, 59(7), pp. 1818-1832, 2014.

9. Enlu Zhou*, Shalabh Bhatnagar, Xi Chen, "Simulation Optimization via Gradient-based Stochastic Search", in Proceedings of the 2014 Winter Simulation Conference, pp. 3869-3879, 2014.

10. Fan Ye and Enlu Zhou*, Dual Formulation of Controlled Markov Diffusions and Its Application", in Proceedings of the 19th IFAC World Congress, pp. 7811-7818, 2014.

New discoveries, inventions, or patent disclosures:

Do you have any discoveries, inventions, or patent disclosures to report for this period?

No

Please describe and include any notable dates

Do you plan to pursue a claim for personal or organizational intellectual property?

Changes in research objectives (if any):

Change in AFOSR Program Officer, if any:

Extensions granted or milestones slipped, if any:

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Jan 29, 2017 22:20:58 Success: Email Sent to: enlu.zhou@isye.gatech.edu